Requested Patent: JP10293755A

Title:

INCIDENCE GRAPH BASED COMMUNICATIONS AND OPERATIONS METHOD AND APPARATUS FOR PARALLEL PROCESSING ARCHITECTURE ;

Abstracted Patent: US6108340 ;

Publication Date: 2000-08-22 ;

Inventor(s): ROLFE DAVID B (US); WACK ANDREW P (US) ;

Applicant(s): IBM (US) ;

Application Number: US19970821894 19970321 ;

Priority Number(s): US19970821894 19970321 ;

IPC Classification: H04L12/42 ;

Equivalents:

ABSTRACT:

A method and apparatus for passing messages between nodes in a distributed network of interconnected nodes wherein a two dimensional array is arranged with each of the nodes represented by a single row heading and a single column heading. The intersections of row and column headings between which messages are to pass may be provided with a token indicative of this condition. The token may further be associated with message parameters defining the passage of the message and operations to be performed thereon, between the two nodes represented by the intersecting row and column headings. Successive versions of the two dimensional array may be provided to form a three dimensional array for passing messages between nodes over the network via successive communication patterns defined by the successive versions of the two dimensional array.

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平10－293755

(43)公開日　平成10年(1998)11月４日

(51)Int.Cl.[6]		識別記号	FI		
Ｇ０６Ｆ	15/163		Ｇ０６Ｆ	15/16	３１０Ｊ
	13/00	３５１		13/00	３５１Ａ

審査請求　未請求　請求項の数29　ＯＬ　（全 17 頁）

(21)出願番号　特願平10－51642

(22)出願日　平成10年(1998)３月４日

(31)優先権主張番号　０８／８２１８９４

(32)優先日　1997年３月21日

(33)優先権主張国　米国（ＵＳ）

(71)出願人　390009531
インターナショナル・ビジネス・マシーンズ・コーポレイション
ＩＮＴＥＲＮＡＴＩＯＮＡＬ　ＢＵＳＩＮＥＳＳ　ＭＡＳＣＨＩＮＥＳ　ＣＯＲＰＯＲＡＴＩＯＮ
アメリカ合衆国10504、ニューヨーク州　アーモンク　（番地なし）

(72)発明者　デビッド・ロルフ
アメリカ合衆国12491、ニューヨーク州ウエスト・ハーレイ、バイン・ツリー・ロード　24

(74)代理人　弁理士　坂口　博　（外１名）

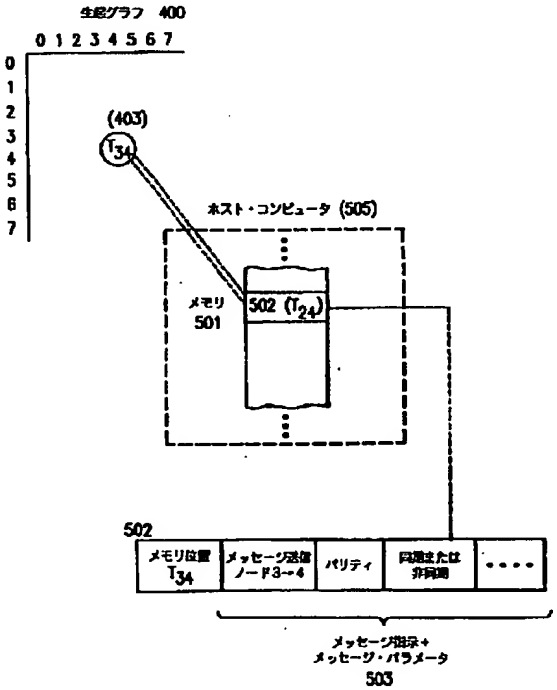
最終頁に続く

(54)【発明の名称】　ネットワーク内のノード間でメッセージを転送する方法

(57)【要約】

【課題】　ノードが相互接続されるネットワークを通じて、ノード間でメッセージを転送する方法及び装置を提供する。

【解決手段】　ノードが相互接続される分散ネットワーク内のノード間で、メッセージを転送する方法及び装置が提供される。２次元配列が、１つの行見出し及び１つの列見出しにより表されるノードにより構成される。メッセージが転送される行及び列見出しの交点に、この条件を示すトークンが提供される。トークンは更に、交差する行及び列見出しにより表される２つのノード間のメッセージの通行、及びそれに対して実行される操作を定義するメッセージ・パラメータに関連付けられる。２次元配列の連続バージョンが提供され、３次元配列を形成し、メッセージが、２次元配列の連続バージョンにより定義される連続通信パターンを介して、ネットワークを通じてノード間で転送される。



【特許請求の範囲】
【請求項１】ネットワーク内のノード間でメッセージを転送する方法であって、前記ノードの各々が前記メッセージの出所または宛先となることができ、前記ノードの各々がメッセージを前記ネットワーク内の任意の他のノードに転送し得るものにおいて、
前記ノードの表現を２次元配列内の列見出し及び行見出しとして構成するステップであって、前記各ノードの表現が前記列見出しとして１度、及び前記行見出しとして１度出現する、ノード表現の構成ステップと、
トークンを前記２次元配列内の前記ノード表現の交点に配置するステップであって、前記トークンが、前記交差ノード表現により表される前記ノード間で、メッセージが転送されるか否かを示す指示に関連付けられ、前記メッセージが転送される場合、前記トークンが、前記ノード間で転送される前記メッセージのメッセージ・パラメータに関連付けられ、こうして生成された前記配列が、前記ネットワークを通じる前記ノード間での前記メッセージの転送のための通信パターンを実現する、トークンの配置ステップと、
前記トークンの各々を前記指示及び前記メッセージ・パラメータに関連付けるステップと、
前記２次元配列内の前記ノード表現の交点における前記トークンが、前記メッセージが転送されることを示す指示に関連付けられる場合、前記トークンを有する前記ノード間で前記メッセージを転送するステップであって、前記メッセージの転送が、前記トークンに関連付けられる前記メッセージ・パラメータにより定義される、メッセージの転送ステップと、
を含む、方法。
【請求項２】前記ノードの各々が、該ノードのために転送される前記メッセージに対応する前記２次元配列の少なくとも一部を記憶する請求項１記載の方法
【請求項３】前記ノードの各々が、前記２次元配列が記憶されるホスト・ノード内のメモリ位置をアクセスできる、請求項１記載の方法。
【請求項４】前記ノードの２つの間で前記メッセージが転送されない場合、２つの前記ノードの前記ノード表現の交点における前記トークンが、前記指示に関連付けられず、前記メッセージ・パラメータに関連付けられない、請求項１記載の方法。
【請求項５】前記２次元配列がコンピュータ・プログラムにより生成される、請求項１記載の方法。
【請求項６】前記ネットワーク内のノードの各々が、前記ノードにより実行される同一の命令セットに従い前記２次元配列を調査し、前記ノードの各々が、前記命令の実行を通じて、前記２次元配列内の前記トークンに従い、前記メッセージを他のノードに転送可能である、請求項１記載の方法。
【請求項７】前記ネットワークがホスト・ノードを含み、該ホスト・ノードが、前記指示、及び前記ネットワークを通じる前記ノード間の前記メッセージの転送を定義する前記メッセージ・パラメータを提供するホスト・ノード・メモリを有し、前記メッセージが転送される前記ノードに対する前記ノード表現の交点に配置される前記トークンが、前記指示及び前記メッセージ・パラメータが記憶される前記ホスト・ノード・メモリのアドレスであるものにおいて、
前記メッセージが転送される前記ノードに対して、前記指示及び及び前記メッセージ・パラメータを、前記２次元配列内の前記ノード表現の交点における前記トークンの各々により示される前記ホスト・ノード・メモリの前記ホスト・ノード・メモリ・アドレスに配置するステップと、
前記ホスト・ノード・メモリから、前記トークンの各々により示される前記ホスト・ノード・メモリ・アドレスに記憶される前記指示及び前記メッセージ・パラメータを検索するステップと、
を含む、請求項１記載の方法。
【請求項８】前記２次元配列内の前記ノード表現の交点に配置される前記トークンが、前記ノードにより復号され得る符号化ワードであり、前記メッセージが前記ノードにより転送されるか否かの前記指示、及び前記メッセージ転送操作を定義する前記トークンに関連付けられる前記メッセージ・パラメータを決定する、請求項１記載の方法。
【請求項９】前記メッセージ・パラメータが、前記メッセージが同期伝送または非同期伝送により転送されるかを示す、請求項１記載の方法。
【請求項１０】前記メッセージ・パラメータが、転送される前記メッセージのビット長を含む、請求項１記載の方法。
【請求項１１】前記メッセージ・パラメータが、前記メッセージの伝送が計時されるか否かを含む、請求項１記載の方法。
【請求項１２】前記ノード間でメッセージを転送するステップが繰返し実行され、前記メッセージの転送を処理する前記ネットワークの能力が、時間の経過と共に低下するか否かを判断する、請求項１記載の方法。
【請求項１３】前記メッセージ・パラメータが、前記メッセージが転送された後に、該メッセージに対して実行される操作を含む、請求項１記載の方法。
【請求項１４】前記２次元配列の前記行見出しとして出現する前記ノード表現が、前記メッセージの転送元の前記ノードを表し、前記２次元配列の前記列見出しとして出現する前記ノード表現が、前記メッセージの転送先の前記ノードを表す、請求項１記載の方法。
【請求項１５】前記２次元配列内の各ノードに対して、前記行見出しとして出現する前記ノード表現の各々と、前記２次元配列内の各ノードに対して、前記列見出しと

して出現する前記ノード表現の各々との間の各交点を調査するステップと、
前記行見出しと前記列見出しとの前記調査済みの交点の各々が、2つの前記ノード間でメッセージが転送されることを示す指示に関連付けられる前記トークンの1つを含むか否かを判断するステップと、
前記調査済みの交点が、2つの前記ノード間でメッセージが転送されることを示す指示に関連付けられる前記トークンの1つを含む場合、前記2次元配列上に前記行見出しとして出現する前記ノード表現により表される前記ノードの各々から、前記2次元配列上に前記列見出しとして出現する前記ノード表現により表される前記ノードの各々へ、メッセージを転送するステップと、
を含む、請求項14記載の方法。
【請求項16】前記調査済みの交点が、2つの前記ノード間でメッセージが転送されることを示す指示に関連付けられる前記トークンの1つを含む場合に、前記2次元配列上に前記行見出しとして出現する前記ノード表現により表される前記ノードから、前記2次元配列上に前記列見出しとして出現する前記ノード表現により表される前記ノードの各々へ、前記メッセージを転送するステップが、前記トークンに関連付けられる前記メッセージ・パラメータに従い実行される、請求項15記載の方法。
【請求項17】前記2次元配列の前記列見出しとして出現する前記ノード表現が、前記メッセージの転送元の前記ノードを表し、前記2次元配列の前記行見出しとして出現する前記ノード表現が、前記メッセージの転送先の前記ノードを表す、請求項1記載の方法。
【請求項18】前記2次元配列内の各ノードに対して、前記列見出しとして出現する前記ノード表現の各々と、前記2次元配列内の各ノードに対して、前記行見出しとして出現する前記ノード表現の各々との間の各交点を調査するステップと、
前記行見出しと前記列見出しとの前記調査済みの交点の各々が、2つの前記ノード間でメッセージが転送されることを示す指示に関連付けられる前記トークンの1つを含むか否かを判断するステップと、
前記調査済みの交点が2つの前記ノード間でメッセージが転送されることを示す指示に関連付けられる前記トークンの1つを含む場合、前記2次元配列上に前記列見出しとして出現する前記ノード表現により表される前記ノードの各々から、前記2次元配列上に前記行見出しとして出現する前記ノード表現により表される前記ノードの各々へ、メッセージを転送するステップと、
を含む、請求項17記載の方法。
【請求項19】前記調査済みの交点が、2つの前記ノード間でメッセージが転送されることを示す指示に関連付けられる前記トークンの1つを含む場合に、前記2次元配列上に前記列見出しとして出現する前記ノード表現により表される前記ノードから、前記2次元配列上に前記行見出しとして出現する前記ノード表現により表される前記ノードの各々へ、前記メッセージを転送するステップが、前記トークンに関連付けられる前記メッセージ・パラメータに従い実行される、請求項18記載の方法。
【請求項20】前記2次元配列の連続バージョンが構成され、前記2次元配列の前記連続バージョンの各々が、異なる前記ノード表現の交点における前記ノード間で前記メッセージが転送されることを示す指示に関連付けられるトークンを含むことができ、前記ネットワークを通じる前記ノード間の前記メッセージの転送のために、連続する潜在的に異なる通信パターンを実現するものにおいて、
前記2次元配列の前記連続バージョンを構成し、3次元配列を形成するステップであって、前記第3次元配列が前記2次元配列の前記連続バージョンを含む、構成ステップと、
トークンを前記2次元配列の前記連続バージョンの各々内の前記ノード表現の交点に配置するステップであって、前記トークンが、前記交差ノード表現により表される前記ノード間でメッセージが転送されるか否かを示す指示に関連付けられ、前記メッセージが転送される場合、前記トークンが、前記ノード間で転送される前記メッセージのメッセージ・パラメータに関連付けられ、こうして生成された前記配列の各々が、前記ネットワークを通じる前記ノード間での前記メッセージの転送のための通信パターンを実現する、トークンの配置ステップと、
前記トークンの各々を前記指示及び前記メッセージ・パラメータに関連付けるステップと、
前記2次元配列の前記連続バージョンの各々の前記ノード表現の交点における前記トークンが、前記メッセージが転送されることを示す指示に関連付けられる場合、前記トークンを有する前記ノード間で前記メッセージを連続的に転送するステップであって、前記メッセージの転送が、前記トークンに関連付けられる前記メッセージ・パラメータにより定義される、メッセージの転送ステップと、
を含む、請求項1記載の方法。
【請求項21】2つの前記ノード間で前記メッセージが転送されない場合、それらの前記ノード表現の交点の前記トークンが0である、請求項20記載の方法。
【請求項22】前記2次元配列の前記連続バージョンの各々が、コンピュータ・プログラムにより生成される、請求項20記載の方法。
【請求項23】前記2次元配列の前記連続バージョンの各々の前記行見出しとして出現する前記ノード表現が、前記メッセージの転送元の前記ノードを表し、前記2次元配列の前記連続バージョンの各々の前記列見出しとして出現する前記ノード表現が、前記メッセージの転送先の前記ノードを表す、請求項20記載の方法。

【請求項２４】前記メッセージを転送するステップが、前記２次元配列の前記連続バージョン上の前記行見出しにより定義される前記ノード表現の各々と、前記２次元配列の前記連続バージョン上の前記列見出しにより定義される前記ノード表現の各々との交点を調査するステップと、
前記ノード表現の前記調査済みの交点の各々に対して、前記メッセージが転送されることを示す指示に関連付けられる前記トークンの１つが、前記交点に含まれるか否かを判断するステップと、
前記調査済みの交点が、２つの前記ノード間でメッセージが転送されることを示す指示に関連付けられる前記トークンの１つを含む場合、前記２次元配列の前記連続バージョン上に前記行見出しとして出現する前記ノード表現により表される前記ノードの各々から、前記２次元配列の前記連続バージョン上に前記列見出しとして出現する前記ノード表現により表される前記ノードの各々へ、前記メッセージを転送するステップと、
を含む、請求項２３記載の方法。
【請求項２５】前記調査済みの交点が、２つの前記ノード間でメッセージが転送されることを示す指示に関連付けられる前記トークンの１つを含む場合に、前記２次元配列の前記連続バージョン上に前記行見出しとして出現する前記ノード表現により表される前記ノードから、前記２次元配列の前記連続バージョン上に前記列見出しとして出現する前記ノード表現により表される前記ノードの各々へ、前記メッセージを転送するステップが、前記トークンに関連付けられる前記メッセージ・パラメータに従い実行される、請求項２４記載の方法。
【請求項２６】前記２次元配列の前記連続バージョンの各々の前記列見出しとして出現する前記ノード表現が、前記メッセージの転送元の前記ノードを表し、前記２次元配列の前記連続バージョンの前記行見出しとして出現する前記ノード表現が、前記メッセージの転送先の前記ノードを表す、請求項２０記載の方法。
【請求項２７】前記メッセージを転送するステップが、前記２次元配列の前記連続バージョン上の前記列見出しにより定義される前記ノード表現の各々と、前記２次元配列の前記連続バージョン上の前記行見出しにより定義される前記ノード表現の各々との交点を調査するステップと、
前記ノード表現の前記調査済みの交点の各々に対して、前記メッセージが転送されることを示す指示に関連付けられる前記トークンの１つが、前記交点に含まれるか否かを判断するステップと、
前記調査済みの交点が、２つの前記ノード間でメッセージが転送されることを示す指示に関連付けられる前記トークンの１つを含む場合、前記２次元配列の前記連続バージョン上に前記列見出しとして出現する前記ノード表現により表される前記ノードの各々から、前記２次元配列の前記連続バージョン上に前記行見出しとして出現する前記ノード表現により表される前記ノードの各々へ、前記メッセージを転送するステップと、
を含む、請求項２６記載の方法。
【請求項２８】前記調査済みの交点が、２つの前記ノード間でメッセージが転送されることを示す指示に関連付けられる前記トークンの１つを含む場合に、前記２次元配列の前記連続バージョン上に前記列見出しとして出現する前記ノード表現により表される前記ノードから、前記２次元配列の前記連続バージョン上に前記行見出しとして出現する前記ノード表現により表される前記ノードの各々へ、前記メッセージを転送するステップが、前記トークンに関連付けられる前記メッセージ・パラメータに従い実行される、請求項２７記載の方法。
【請求項２９】前記２次元配列の前記連続バージョンに対する前記メッセージ転送のために、前記ノードの各々が同一の命令セットを実行する、請求項２０記載の方法。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】本発明は、複数の相互接続処理ユニットまたはノードを含む並列コンピュータ・システムの分野に関し、特に生起グラフを用いて、並列コンピュータ・システムの相互接続ネットワーク上の処理ノード間で、メッセージを転送するために使用される通信パターンを定義する通信方法及び装置に関する。本発明は、並列処理システム内の相互通信されるノードの数に関係無しに、並列処理システムにおいて関心のある任意の通信パターンの実現を可能にする。更に、本発明は生起グラフを用いて、並列コンピュータ・システム内の処理ノードにおいて、他の非通信ベースの操作を駆動することを含む。
【０００２】
【従来の技術】並列処理システムは、処理スピードを向上することにより、大量のデータを管理し、複雑な計算問題を迅速且つ効率的に処理するコンピュータ・システムを構築するための有用なアプローチとして使用されてきた。多数の並列または分散処理システムが知られている。
【０００３】大規模並列処理システムは、しばしば数百乃至数千もの非常に多くの別々の、しかしながら比較的単純なマイクロプロセッサ・ベースの処理要素を含み、これらは通信機構を通じて相互接続される。通信機構は通常、高速パケット・ネットワークを含み、そこでは各処理要素がネットワーク上の別々のノードとして現れる。メッセージはパケットの形式を取り、ネットワーク上でこれらの処理要素間を経路指定され、それらの間の通信を可能にする。各処理要素は通常、別々のマイクロプロセッサ及び関連支援回路を含む。こうした関連支援回路には、ランダム・アクセス・メモリ（ＲＡＭ）など

の記憶回路、及び読出し専用メモリ（ROM）回路及び入出力回路、並びに通信サブシステムなどが含まれる。通信サブシステムは、通信インタフェース及び関連ハードウェア及びソフトウェアを含み、これらは処理要素がネットワークとインタフェースすることを可能にする。通信機構は、処理要素またはノードによる命令の同時または並列実行を可能にする。

【0004】多数の相互接続されるノードを含むこうした並列処理システムでは、ノード間通信（またはノード間の通信パターン）の可能な組み合わせの数が、2^{n*n}として成長する。ここで、nはシステム内の相互間でメッセージを転送可能なノードの数を表す。従って、僅かに8つのノードを有する分散コンピュータ・システムは、2^{64}（または約$1.845*10^{19}$）通りの異なる潜在的な通信パターンを示し得る。

【0005】並列コンピュータ・システムにおいて、あらゆる可能な通信パターンを行使するシステムを実現することは実用的でないので、システムがいつでも、関心のある任意の通信パターンを迅速且つ効率的に行使できるようにすることが望ましい。

【0006】例えば、一般性を失うこと無しに、通常の並列処理システム内で相互接続されるノードをテストするテスト・プログラムの生成について考えてみる。現テスト・プログラムは固定通信パターンを実現し、その結果、並列コンピュータ・システムの通信サブシステムが十分にテストされない。通常の現生成テスト・プログラムは、特定の通信パターンをテストするために生成される。こうしたテスト・プログラムの生成は、労働集約型のプロセスであり、かなりのプログラミング資源を必要とする。従って、並列コンピュータ・システム内で実現され得る様々な通信パターンをテストするための、多数のこうしたテスト・プログラムの生成は、多大な量のプログラミング資源を消費する。結果的に、現テスト技術を鑑みると、分散コンピュータ・システムにおける通信パターンの十分なテスト範囲を達成するテスト・プログラムの作成は、法外に高価な尽力を必要とする。

【0007】抽象的2項関係を定義する図形解析の利用が、従来理解されてきた。並列コンピュータ・システム・トポロジが、高度な故障許容を達成する図形解析技術を用いて構築された。例えば、Bruckらによる米国特許番号第5280607号、"Method and Apparatus for Tolerating Faults in Mesh Architectures"（1994年1月18日発行）は、並列コンピュータ・システムにおいて使用される故障許容メッシュ・アーキテクチャの構築について述べている。この発明は、図形構築を通じて実現され、円グラフが同一の要素を表し、グラフの縁が、所与のターゲット・メッシュに対して設計される故障許容メッシュのノード間の通信リンクを表す。所定数の可能な故障に対して、このように設計される故障許容メッシュが、サブグラフとして、ターゲット・メッシュに対応するグラフを含むように保証される。

【0008】図形解析は更に、並列コンピュータ・システム内の処理ノード間の改善されたメッセージ・ルーティングを実現するために使用された。Petersonらによる米国特許番号第5170393号、"Adaptive Routing of Messages in Parallel and Distributed Processor Systems"は、ヒューリスティック・ベースの適応ルーティング・アルゴリズムを実現することにより、通信ノード間のメッセージ待ち時間を低減する方法及び装置について述べている。これは障害のある通信経路セグメントを考慮から除外することにより、ノード間通信を最小時間で成功または失敗に導く。Petersonらにより述べられるルーティング・モデルは、システムの区分化相互接続グラフを用いて、出所ノードと宛先ノード間の成功経路のために、ネットワークの余分な部分を削除しながら、訪問されなければならない特定のノード・セットを定義する。このプルーニング・アルゴリズムは隣接マトリックス上で実現され、これは1つのホップ（または相互接続グラフの1つの縁）により分離されるノードの交点に、1つを含む。

【0009】Wangらによる論文"Scheduling of Unstructured Communications on the Intel iPSC/860"（Supercomputing '94、Proceedings、p.360以降参照、1994年）では、高度並列コンピュータ・システムの所定の通信パターンを表す通信マトリックスが、多数のアルゴリズムを通じて、部分順列に分解される。これらの部分順列は、高度並列コンピュータ・システムにおいて、ノード及びリンク競合を回避するように、全部対複数（all-to-many）の個人化通信をスケジュールするために使用される。

【0010】Wangらの論文の状況では、通信マトリックスがスケジューリング・アルゴリズムの入力として使用され、この発明は競合を回避するために、既知の通信パターンに対して、メッセージをスケジューリングする。従ってWangらは、既存のプログラムの実行を最適化するために、生起グラフ形式により、そのプログラムに対する通信パターンの獲得について述べている。

【0011】Rolfeによる米国特許番号第5313645号、"Method for Interconnectingand System of Interconnected Processing Elements by Controlling NetworkDensity"（本願の出願人に権利譲渡済み）は、とりわけ、ネットワーク径に対する1要素当たりの接続数を平均化するために、処理要素を相互接続する方法について述べている。この特許は、新規の相互接続技術を達成するネットワーク接続アルゴリズムの使用を示す。大規模並列コンピュータ・ネットワークにおける物理ノード相互接続（すなわちケーブル・リンクまたは他の手段を介する）の定義に論点が向けられているが、このアルゴリズムは、並列コンピュータ・システム内で多数のこうしたパターンを行使するための、論理通信経路の生成に

も適応され得る。従って、重要な通信パターンのセットに対応する生起グラフの生成が、単一のコンピュータ・プログラムにより実現されるこのアルゴリズムを通じて生成され得る。そして生起グラフは、超立方体、及び相互接続される円弧を含む直径可変のｎ次元トーラスを含み得る。

【００１２】並列処理ネットワークにおいて、この自動的に生成される通信パターンのセットを入力として使用することにより、これらの生起グラフ（または通信マトリックス）が、並列コンピュータ・システム内のノード間のメッセージの進行を方向付ける役割をするシステムを実現することが可能となり、並列コンピュータ・システムにとって、特に関心のある任意の通信パターンを実現できる。更に、こうしたシステムでは、並列コンピュータ・システム内の処理ノードにおいて、非通信ベースの操作の実行を駆動するためにも、生起グラフが代用され得る。

【００１３】

【発明が解決しようとする課題】本発明は、ノードが相互接続されるネットワークを通じて、ノード間でメッセージを転送する方法及び装置を提供するものであり、従来技術の前述の問題及び欠点を克服し、更に追加の利点を提供する。

【００１４】

【課題を解決するための手段】第１の実施例では、２次元配列が、各ノードが行見出しとして１度、及び列見出しとして１度出現する表現により構成される。配列内において、行及び列ノード表現の交点が、各々トークンを提供され、トークンは、交差ノード表現により表されるノード間でメッセージの転送が発生するか否かの指示に関連付けられる。メッセージが、２次元配列内の交差ノード表現により表される２つのノード間で転送される場合、トークンは更に、転送されるメッセージのタイプ、及びメッセージの転送方法を管理するメッセージ・パラメータに関連付けられる。従って、２次元配列は、ネットワーク上のノードにより行使される通信パターンを含む。各ノードは配列内容を考慮し、自身が他のどちらのノードにメッセージを送信するか、及び自身が他のどちらのノードからメッセージを受信するかを判断する。従って、メッセージは配列内のトークン表現に従い、ノード間を転送される。

【００１５】別の実施例では、ネットワークを通じるノード間でのメッセージの転送が、ホスト・ノードの参照を介して達成され、ホスト・ノードがホスト・ノード・メモリを有する。この実施例では、ノード表現の交点におけるトークンのサブセットが、０の形式を取ることができ、これはメッセージが、２次元配列内のノード表現の交点により表されるノード間で転送されないことを示す。配列内の他のトークンは、ホスト・ノード・メモリのアドレスに関連付けられ得る。これらのトークンは、メッセージが前記交点により表されるノード間で転送されることを示す役割をする。これらの交点において、トークンに関連付けられるアドレスにより示されるホスト・ノード・メモリ位置は、ノード間を転送されるメッセージのタイプ、及びメッセージが転送される方法を示すメッセージ・パラメータを記憶し、これらには例えば、受信ノードにおいて受信されるメッセージ・データに対して実行される操作などが含まれる。

【００１６】更に別の実施例では、前述の２次元配列の連続バージョンが、３次元配列を形成するように構成され、第３次元が２次元配列の連続バージョンを含む。２次元配列の各連続バージョンにおいて、ノード表現の各交点におけるトークンが、異なる指示に関連付けられ、それにより各配列内のノードのネットワークに対して、異なる通信パターンを含む。この実施例では、各２次元配列内で実現される通信パターンが、ノードが相互接続されるネットワーク内のノードにより、単一の命令セットを通じて実現され、それによりシステムが適用され得る使用範囲を拡大する。

【００１７】従って、本発明によれば、コンピュータ・プログラムなどの既知の手段を通じて生成され得る配列が、並列処理システムにおいて関心のある任意の通信パターンを実現するために使用され得る。このように、処理ノードの並列ネットワークの拡張利用が達成され得る。

【００１８】本発明の更に別のアプリケーションでは、各ノードが行見出しとして１度、及び列見出しとして１度出現する表現により、２次元配列が構成され得る。トークンが行列交差において提供され、算術または診断操作などの非通信ベースの操作がそこで実行されるか否かに関する指示に関連付けられる。トークンは更に、操作パラメータにも関連付けられ、これらには実行される操作のための命令、及び操作を完了に運ぶために必要なあらゆるオペランドが含まれる。従って、この代替実施例では、２次元配列がネットワークの操作マトリックスを表し、そこでは各ノードが配列を考慮して、自身が実行する操作を決定する。

【００１９】

【発明の実施の形態】本発明の原理によれば、ノードのネットワークを通じて、ノード間でメッセージを転送する方法及び装置が開示され、これは特に、大規模並列コンピュータ・システム内のコンピュータ要素間において当てはまる。本発明について詳述する前に、幾つかの典型的な通信パターンを通じて、本発明の方法及び関連装置により実現され得る有効性について述べることが有益であろう。

【００２０】図１を参照すると、８つの処理ノード（０乃至７）のセットにおける単純な環状通信パターンが示される。ここで重要な点は、図示のノード間の通信が、論理通信またはいわゆる通信トポロジ若しくは通信パタ

ーンを表しており、相互接続されるノードの物理トポロジを表すものでないということである。図示の通信パターンでは、各ノードはメッセージを最も近い隣接ノードに送信するように、またそれらのノードからメッセージを受信するように指示され、このことが2重矢印の実線で示される。従って、例えば図1において、ノード1はメッセージをノード0及びノード2に送信し、同様にメッセージをノード0及びノード2から受信する。並列処理システムにおいて、この通信パターンを実現するために、特定のアプリケーション・プログラムを生成することが必要であろう。同一の並列処理システムにおいて、多少異なる通信パターンが関心となる場合には、この第2のパターンを実現する別のアプリケーション・プログラムを生成することが必要となろう。例えば、図1において2重矢印の破線で示される環状通信トポロジ、すなわち各ノードがメッセージを、その最も近い隣接ノード及び1つ置きのノードに送信するトポロジでは（例えば図1において、ノード1がメッセージをノード0及び2、並びにノード7及び3に送信する）、最初の典型的で単純な環状通信トポロジを実現するために生成されるプログラムとは異なる別の通信プログラムを独立に生成することが要求される。

【0021】図2を参照すると、2次元トーラスの通信パターンが示される。2次元トーラスでは、図示の通信トポロジ内の各ノードが、それに最も近い4つの隣接ノードと通信する。例えば、ノードA1はノードA2、A3、A4及びA5と通信する。従って、並列処理システム内のノード間で、このようにメッセージが転送されるようにするために、この通信パターンを実現する別の通信プログラムを生成することが要求される。

【0022】考慮される最後の例として、図3は5つのノードにおける、全組み合わせの通信パターンを示す。例えば、ノード1は図示のように、ノード2、3、4及び5の任意のノードと通信する。再度、この通信パターンは、並列処理システムの通信機構において、これらのノード間通信を実現するための別のプログラムの生成を要求する。

【0023】前述の説明から、個別の通信パターンの実現は、所与の並列処理システムにおいて考慮される各パターンに付随し、別々のアプリケーション・プログラムの生成を要求することが理解される。並列プログラムの作成は時間集約型の尽力を必要とし、n個のノードを有するコンピュータ・ネットワーク内には、考慮され得る2^{n*n}の別個のパターンが存在するので、特定のネットワーク内で実現される全ての通信パターンに対して、別々のプログラムを個別に生成することは非現実的であることは明らかである。従って、関心のある任意の通信パターンが、1つの並列プログラムを通じて実現され得るように、並列処理システム内の処理ノードにより実行されるプログラムを、効率的に生成することが待望されている。こうしたニーズは従来技術では十分に解決されなかった。

【0024】本発明の1つの態様は、この未解決のニーズに関わるものである。本発明の特定の態様によれば、複数のコンピュータ・ノード間で、設計者にとって関心のある任意の通信パターンを実現する1つのアプリケーション・プログラムを生成する方法及び関連装置が開示される。操作において、アプリケーション・プログラムの実行により、通信パターンを定義する役割をする生起グラフのセットが生成され得る。処理ノードに対応する行見出し及び列見出しを有する生起グラフを参照することにより、処理ノードの各々がイネーブルされ、メッセージが転送される2つのノードに対応する生起グラフ内の交点に含まれるトークン従い、メッセージを並列処理システム内の他の処理ノードに転送する。

【0025】次に図4を参照しながら、本発明の、より詳細について述べることにする。図4は、8つのノードを有し、図1に示される単純な環状通信パターン100を実現するネットワークの2次元生起グラフ400の単純な例を示す。行見出し及び列見出しの各々は、並列処理システム内の処理ノードを表す。一般性を欠くこと無しに、行見出しを含むノード表現は、"送信ノード"（401_0乃至401_7）として指定され、各添字は、並列処理システム内の異なる"送信"ノードを指定する。送信ノード0乃至7は、所与の2つのノード間で転送されるメッセージの出所として作用する。同様に、列見出しを含むノード表現が、"受信ノード"（402_0乃至402_7）として指定される。受信ノードは、所与の2つのノード間を転送されるメッセージが移動する宛先として作用するノードである。典型的なシステムでは、所与のノード（例えばノード1）は、特定のメッセージの通行における送信ノードとして作用し（例えばノード1がメッセージをノード2に送信する）、また特定のメッセージの受信のための受信ノードとしても作用する（例えばノード1がメッセージをノード2から受信する）。従って、送信ノードとしてのノード1　401_1に対応する行表現は、典型的な生起グラフ400内において、受信ノードとしてのノード1　402_1に対応する列表現を伴う。

【0026】図4に示される生起グラフ400は、多数の既知のコンピュータ・ベースの実施例により生成され得る。例えば、Rolfeによる米国特許番号第5313645号、"Method for Interconnecting and System of Interconnected ProcessingElements by Controlling Network Density"では、ネットワーク密度を制御しながら、ネットワーク内のノード間の物理的相互接続を確立するネットワーク接続アルゴリズムが開示される。この特許は、分散ネットワーク上での物理ノードの相互接続を定義するコンピュータ・プログラムを実現可能なアルゴリズムを示しているが、これは図示の生起グラフ400などの、並列コンピュータ・システムの分散ノード間

のメッセージ・ルーティングを示す配列の生成にも、同様に適用可能である。生起グラフ４００などの典型的な生起グラフを生成するための、他の方法も考案され得り、これらについても本発明の範囲内に含まれるものと見なされる。

【００２７】図4から分かるように、生起グラフ内の特定の行及び列の交点は、トークン"T"４０３を含む。トークンを含むことは、メッセージが行見出し及び列見出しにより表されるノード間で転送されることを示す。例えば、図４では、行３と列４との交点に示されるトークン"T_{34}"４０３は、メッセージが並列処理システム内で、処理ノード３から処理ノード４に転送されることに対応する。この図示の例において含まれるトークンは、行見出しと列見出しとの交点に示されるように単純であり、これは単にメッセージがその行見出しに対応する処理ノードから、その列見出しに対応する処理ノードに転送されることを意味する。従って、行見出しにより表される処理ノード（または送信ノード）（この例におけるノード3401_3）は、送信操作を実行し、列見出しにより表される処理ノード（または受信ノード）（この例におけるノード４　402_4）は、対応する受信操作を実行し、それにより並列システム通信ネットワークを通じるメッセージの通行を可能にする。所与の行見出し及び列見出しの交点に"０"が示される場合、その行見出しに対応する処理ノードから、その列見出しに対応する処理ノードに、メッセージは転送されない。ここで、生起グラフ４００の特定の行見出し及び列見出しの交点に、特定のトークンを含む場合は、２つの表されるノード間のメッセージの通行のための命令に対応し、一方、そこに異なるトークンを含むか、トークンを含まない場合には、それらの間でメッセージの通行が発生しない命令に対応することが理解されよう。

【００２８】別の、より高性能な実施例では、トークンを含むことが、交差する行と列とに対応する処理ノード間の通信操作を、より厳格に定義するために使用され得る。例えば、特定の行及び列見出しの交点に示されるトークンが、表される処理ノード間で転送されるメッセージのパラメータに関連付けられ、これらのパラメータには、転送されるメッセージのビット長やパリティなどが含まれる。更に、または代わりに、メッセージが転送される通信プロトコルにトークンが関連付けられてもよく、それらには、例えばメッセージが同期または非同期伝送のいずれにより転送されるのか、またはメッセージの伝送が計時されるか否か、または通信操作を制御するために使用され得る任意の他の特性などが含まれる。更に、トークンが、ノード間メッセージ内で転送されるデータに対して、送信ノードまたは受信ノードのいずれかで実行される操作セットまたは命令セットに関連付けられ得る。それらには、例えば送信データと比較して、受信ノードにおいて受信されるデータの完全性を評価する巡回冗長検査（ＣＲＣ）、受信ノードにおいて受信されるデータにもとづく計算、または送信ノードからの伝送以前に、データに対して実行される計算などが含まれる。更に、トークンと命令セットとの関連付けなどにより、単一命令複数データ・ストリーム（ＳＩＭＤ）または複数命令複数データ・ストリーム（ＭＩＭＤ）のいずれかの並列コンピュータ・プログラムが、対応する生起グラフにもとづき実現され得る。

【００２９】上述のメッセージ転送の実現は、ネットワーク・メッセージ転送能力の低下を時間の経過と共にテストする目的で、並列コンピュータ・システム内で繰り返し実現され得る。こうしたテストを規定するために、ノードのネットワークにとって特に有用な通信パターンを定義する典型的な生起グラフ４００が、ノードにより繰り返し実現され、続く通信の結果がモニタされ、ネットワーク上のルーティング性能が時間の経過と共に変化するか否かに注目する。

【００３０】操作において、生起グラフ４００の行及び列の交点に配置される前述のトークンは、前述のようなメッセージ通行プロトコルに関する情報が記憶されるコンピュータ・メモリ内のアドレス、または特殊目的ハードウェア・レジスタ若しくはマイクロコード位置などを参照する。或いは、トークンが、符号化された２進ワードまたは１６進ワードの形式を取り、ハードウェアまたはソフトウェア手段を介する復号の際に、前述のような通信プロトコルの特定の面を示す値を生成してもよい。トークン化インタフェースの更に別の例に関しては、本願の出願人に権利譲渡された米国特許番号第５４８５６２６号、"Architectural Enhancementsfor Parallel Computer Systems Utilizing Encapsulation of Queuing Allowing Small Grain Processing"（１９９６年1月１６日発行）を参照されたい。例えば、図５を参照すると、本発明の実施例が示され、トークンT_{34}　４０３などの各トークンが、ホスト・コンピュータ５０５のメモリ５０１内の特定のメモリ位置５０２を参照する。ホスト・コンピュータ５０５は、生起グラフ４００上の行４０１及び列４０２の見出しにより表される、並列処理システム内の複数の処理ノード（すなわちノード０乃至ノード７）の１つである。特定のトークン４０３により参照されるホスト・コンピュータ・メモリ・アドレス５０２は、交差ノードのノード間通信操作を適切に制御するために必要なデータを含む。より一般的には、図６を参照すると、生起グラフ４００を形成するトークン６００が、メモリ・アドレス６０２の参照を通じて、または復号プロセス６０３を通じて、メッセージが２つの交差ノード間で転送されるか否かを決定する指示６０４に関連付けられる。メッセージが転送される場合、トークンは更に、通信操作及び続く任意の操作の実行ために必要なメッセージ・パラメータ６０６（例えばパリティ、メッセージ転送の計時の有無、受信時のＣＲＣなど）に関連

付けられる（６０５）。

【００３１】ここで述べられる生起グラフ・ベースの通信方法を実現するために、生起グラフ４００上に表されるノードの各々は、生起グラフを調査し、それが実行しなければならない操作を見分けなければならない。典型的な実施例では、生起グラフ４００上で送信ノードとして表される各ノードは、各受信ノードとの交点を調査する。図７のフロー図は、ネットワーク内の全てのノードに対して、典型的な生起グラフ４００を用いて、意図した通信パターンの行使を可能にするステップを示す。生起グラフ４００上の各ノードの表現に対して、生起グラフの交点が調査され、そのノードが他のどちらのノードにメッセージを送信するかが決定される。従って、図７は、一般性を欠くこと無しに、行見出しが送信ノードとして、列見出しが受信ノードとして定義される生起グラフ４００に当てはまる。第１のステップ７０１では、第１のノード（例えばノード０　401_0）が選択される。選択ノードに対して、そのノードを見出しとして有する行を横切る第１の交点が調査される（ステップ７０２）。メッセージが送信されることを示すトークンが存在する場合（ステップ７０３）、ノードはステップ７０４に移行し、トークンとメッセージ指示６０４及び関連メッセージ・パラメータ６０６との関連付けを解読し（図６の６０２または６０３）、次に送信操作を実行し（ステップ７０５）、交差する列見出しにより表され、受信操作を実行するように命令されるノードに、メッセージを送信する。次にステップ７０６で、調査済みの交点が、所与の行内の最後の交点であるか否かが決定されなければならない。他方、ステップ７０３で、トークンがメッセージの送信指示に関連付けられない場合、実行はステップ７０６に直接移行する。ステップ７０６で、調査中の所与の行内に更に交点が存在すると判断される場合、実行はステップ７０２に戻り、次の交点が調査される。代わりに、所与の行の最後の交点が調査された場合には、実行はステップ７０７に移行し、生起グラフ４００上の行見出しを有する全てのノードが、評価されたか否かが判断される。否定の場合、次のノード（例えばノード１　401_1）が選択され（ステップ７０８）、実行は新たに選択された行の交点の調査に戻る（ステップ７０２）。最後のノードが調査されたとき、操作はその完了に達する（ステップ７０９）。もちろん、生起グラフ４００の列方向の調査においても、同一のタイプの操作が実行され、トークンが列の見出しのノードに受信操作を実行するように命令し、関連付けられる送信ノードが送信操作を実行するように命令される。更に、生起グラフの別の実施例では、列及び行見出しが送信ノード及び受信ノードにそれぞれ対応し、前述のステップの実行がそれに応じて進行する。図示のフロー図は、異なる生起グラフ及び特定の生起グラフの異なる実施例に容易に適応され、全てが本発明の範囲に含まれるものと見なされる。

【００３２】図７に示される生起グラフ４００の順次的な調査は、ノードのネットワーク上で実行される並列コンピュータ・プログラムを通じて実現されることができ、ネットワーク内の各ノード（例えば生起グラフ４００では、ノード０乃至７）は、各処理ノードのメモリに記憶される生起グラフ４００のコピーを提供されたり、或いは生起グラフ４００のバージョンが、例えばネットワーク内の各処理ノードにより読出されるホスト・ノード５０５のメモリ５０１などの、共通にアクセスされるメモリ位置に保持されたり、或いは特定のノードにとって関心のある生起グラフ部分だけが、そのノードに提供されたりアクセス可能である。例えば、生起グラフ４００上のノード０の場合、これはベクトル４０４及び４０５を含む生起グラフ４００の部分を、ホスト・ノード・メモリを介してアクセスするか、或いはそれ自身のメモリに記憶する。ここでベクトル４０４は、生起グラフ４００上で401_0を見出しとする行の列交点を含み、ベクトル４０５は、生起グラフ４００上で402_0を見出しとする列の行交点を含む。いずれの場合にも、ネットワーク内の各ノードが、少なくともそのノードにおける生起グラフの関連部分へのアクセスを有するか、または記憶する限り、各ノードは、同一の命令セットの実行を通じてネットワーク内の他のノードの各々に対する通信パターンとは異なる、特定の通信操作に従事する。例えば、ノード１は、生起グラフ４００へのアクセスを通じて（または生起グラフ４００の少なくとも行１及び列１へのアクセスを通じて）、命令セットを実行し、生起グラフ内で決定されるメッセージの通行（すなわち、ノード０及びノード２との間のメッセージの通行）を可能にする。一方、同一の生起グラフ４００へのアクセス及び同一の命令セットを通じて、ノード２は、生起グラフ４００の内容により指定される異なる通信操作（すなわち、ノード１及びノード３との間のメッセージの通行）を規定することを許可される。

【００３３】前述の米国特許番号第５３１３６４５号で述べられる典型的なアルゴリズムなどの、コンピュータにより実現されるアルゴリズムにより、典型的な生起グラフ４００を生成する能力が、並列コンピュータ・システムのユーザにとって関心のある異なる通信パターンを含む、非常に多数の生起グラフの生成を容易にする。図８は、３次元配列８００を形成する２次元生起グラフのセットを示し、図４に示される生起グラフ４００などの、２次元生起グラフの連続バージョン８０１乃至８０７が、３次元配列８００の第３次元を形成する。実際には、単一の生起グラフ４００に関し述べたときと同様に、多数の生起グラフが、多数の重要な通信パターンをノードのネットワークに対して、連続的にテストするために使用される。更に、各処理ノードにおいて実行される命令セットが生起グラフの内容に関係無しに同一に維

持され、その結果、複数の生起グラフが、同一のメッセージ転送プログラムへの入力として使用され、それによりプログラムがネットワークに対する異なる通信パターンを生成する。

【００３４】別の実施例では、生起グラフが、非通信ベースの操作が各ノードにおいて実行されるか否かの指示に関連付けられるトークンにより形成される。従って、典型的な生起グラフ４００は、ノード１及び２の交点（T_{12}）にトークンを含み、これが交差ノードに高速フーリエ変換、バス診断、またはノード間のメッセージの転送に関連しない他の操作を実行するように命令する。図9に示されるように、トークン９００が、前述のメモリ・アドレス参照９０２または復号プロセス９０３といったトークン関連付け機構を通じて、交差ノードにおいて操作が実行されるか否かに関する指示９０４に関連付けられ（９０１）、更に、操作の実行のために要求され得るオペランドなどの、任意の操作パラメータ９０６に関連付けられる（９０５）。このように、生起グラフ４００は、各交点における特定のトークンの関連付けにもとづき、メッセージ転送の他に、各ノードにおける個別の非通信ベースの操作を可能にするためにも使用され得る。更に、２つの交差ノード間で、メッセージ転送及び他の操作の実行が所望される場合、これは対応するトークンを複数の通信及び非通信操作に関連付けることにより達成され得る。

【００３５】実際には、複数ノード・ネットワーク内のノードにおいて、操作を可能にする生起グラフベースの調査は、ネットワーク内のノード間でのメッセージの通信に対する調査と類似に、順次的に進行する。図１０は、生起グラフを用いてネットワーク内での操作を可能にするために、ネットワーク内の各ノードにおいて要求されるステップのシーケンスを示す典型的なフロー図である。最初のステップ１００１では、第１のノード（例えばノード０　401_0）が選択される。選択ノードに対して、そのノードを見出しとして有する行を横切る第１の交点が調査される（ステップ１００２）。ある操作が実行されることを示すトークンが存在する場合（ステップ１００３）、ノードはステップ１００４に移行し、トークンと操作指示９０４及び関連操作パラメータ９０６との関連付けを解読し（図９の９０２または９０３）、ノードは操作を実行し（ステップ１００５）、交差する列見出しにより表されるノードも、トークンに関連付けられた操作パラメータ９０６に依存して、同一のまたは異なる操作を実行するように命令され得る。次にステップ１００６で、調査済みの交点が、所与の行内の最後の交点であるか否かが決定されなければならない。他方、ステップ１００３で、トークンがある操作の実行指示に関連付けられない場合、実行はステップ１００６に直接移行する。ステップ１００６で、調査中の所与の行内に更に交点が存在すると判断される場合、実行はステップ１００２に戻り、次の交点が調査される。代わりに、所与の行の最後の交点が調査された場合には、実行はステップ１００７に移行し、生起グラフ４００上の行見出しを有する全てのノードが評価されたか否かが判断される。否定の場合、次のノード（例えばノード１　401_1）が選択され（ステップ１００８）、実行は新たに選択された行の交点の調査に戻る（ステップ１００２）。最後のノードが調査されたとき、操作はその完了に達する（ステップ１００９）。もちろん、生起グラフ４００の列方向の調査においても、同一のタイプの操作が実行され、トークンが列の見出しのノード、及び交差行の見出しのノードに、指示された操作を実行するように命令する。

【００３６】通信ベースの操作において前述したときと同様に、非通信ベースの操作に対する生起グラフ４００の順次評価が、ノードのネットワーク上で実行される並列コンピュータ・プログラムにより実現され得る。ここでネットワーク内のノード（すなわち生起グラフ４００によれば、ノード０乃至７）は、生起グラフ４００のコピーを提供されるか、或いは生起グラフ４００のバージョンが、例えばネットワーク内の各処理ノードにより読出されるホスト・ノード５０５のメモリ５０１などの、共通にアクセスされるメモリ位置に保持されたり、或いは特定のノードとって関心のある生起グラフ部分だけがそのノードに提供されたり、アクセス可能である。例えば、生起グラフ４００上のノード０の場合、これはベクトル４０４及び４０５を含む生起グラフ４００の部分を、ホスト・ノード・メモリを介してアクセスするか、或いはそれ自身のメモリに記憶する。ここでベクトル４０４は、生起グラフ４００上で401_0を見出しとする行の列交点を含み、ベクトル４０５は、生起グラフ４００上で402_0を見出しとする列の行交点を含む。いずれの場合にも、ネットワーク内の各ノードが、その特定ノードにおける生起グラフ４００の関連部分へのアクセスを有するか、または記憶する限り、各ノードは同一の命令セットの実行を通じて、関連トークンにより指示される特定の非通信操作を実行し得る。

【００３７】以上、好適な実施例について詳述してきたが、当業者には、本発明の趣旨から逸脱すること無しに、様々な変更、追加、改善及び拡張が可能であり、これらについても本発明の範囲に含まれることが理解されよう。

【００３８】まとめとして、本発明の構成に関して以下の事項を開示する。

【００３９】（１）ネットワーク内のノード間でメッセージを転送する方法であって、前記ノードの各々が前記メッセージの出所または宛先となることができ、前記ノードの各々がメッセージを前記ネットワーク内の任意の他のノードに転送し得るものにおいて、前記ノードの表現を２次元配列内の列見出し及び行見出しとして構成す

るステップであって、前記各ノードの表現が前記列見出しとして1度、及び前記行見出しとして1度出現する、ノード表現の構成ステップと、トークンを前記2次元配列内の前記ノード表現の交点に配置するステップであって、前記トークンが、前記交差ノード表現により表される前記ノード間で、メッセージが転送されるか否かを示す指示に関連付けられ、前記メッセージが転送される場合、前記トークンが、前記ノード間で転送される前記メッセージのメッセージ・パラメータに関連付けられ、こうして生成された前記配列が、前記ネットワークを通じる前記ノード間での前記メッセージの転送のための通信パターンを実現する、トークンの配置ステップと、前記トークンの各々を前記指示及び前記メッセージ・パラメータに関連付けるステップと、前記2次元配列内の前記ノード表現の交点における前記トークンが、前記メッセージが転送されることを示す指示に関連付けられる場合、前記トークンを有する前記ノード間で前記メッセージを転送するステップであって、前記メッセージの転送が、前記トークンに関連付けられる前記メッセージ・パラメータにより定義される、メッセージの転送ステップと、を含む、方法。

（2）前記ノードの各々が、該ノードのために転送される前記メッセージに対応する前記2次元配列の少なくとも一部を記憶する前記（1）記載の方法

（3）前記ノードの各々が、前記2次元配列が記憶されるホスト・ノード内のメモリ位置をアクセスできる、前記（1）記載の方法。

（4）前記ノードの2つの間で前記メッセージが転送されない場合、2つの前記ノードの前記ノード表現の交点における前記トークンが、前記指示に関連付けられず、前記メッセージ・パラメータに関連付けられない、前記（1）記載の方法。

（5）前記2次元配列がコンピュータ・プログラムにより生成される、前記（1）記載の方法。

（6）前記ネットワーク内のノードの各々が、前記ノードにより実行される同一の命令セットに従い前記2次元配列を調査し、前記ノードの各々が、前記命令の実行を通じて、前記2次元配列内の前記トークンに従い、前記メッセージを他のノードに転送可能である、前記（1）記載の方法。

（7）前記ネットワークがホスト・ノードを含み、該ホスト・ノードが、前記指示、及び前記ネットワークを通じる前記ノード間の前記メッセージの転送を定義する前記メッセージ・パラメータを提供するホスト・ノード・メモリを有し、前記メッセージが転送される前記ノードに対する前記ノード表現の交点に配置される前記トークンが、前記指示及び前記メッセージ・パラメータが記憶される前記ホスト・ノード・メモリのアドレスであるものにおいて、前記メッセージが転送される前記ノードに対して、前記指示及び及び前記メッセージ・パラメータを、前記2次元配列内の前記ノード表現の交点における前記トークンの各々により示される前記ホスト・ノード・メモリの前記ホスト・ノード・メモリ・アドレスに配置するステップと、前記ホスト・ノード・メモリから、前記トークンの各々により示される前記ホスト・ノード・メモリ・アドレスに記憶される前記指示及び前記メッセージ・パラメータを検索するステップと、を含む、前記（1）記載の方法。

（8）前記2次元配列内の前記ノード表現の交点に配置される前記トークンが、前記ノードにより復号され得る符号化ワードであり、前記メッセージが前記ノードにより転送されるか否かの前記指示、及び前記メッセージ転送操作を定義する前記トークンに関連付けられる前記メッセージ・パラメータを決定する、前記（1）記載の方法。

（9）前記メッセージ・パラメータが、前記メッセージが同期伝送または非同期伝送により転送されるかを示す、前記（1）記載の方法。

（10）前記メッセージ・パラメータが、転送される前記メッセージのビット長を含む、前記（1）記載の方法。

（11）前記メッセージ・パラメータが、前記メッセージの伝送が計時されるか否かを含む、前記（1）記載の方法。

（12）前記ノード間でメッセージを転送するステップが繰返し実行され、前記メッセージの転送を処理する前記ネットワークの能力が、時間の経過と共に低下するか否かを判断する、前記（1）記載の方法。

（13）前記メッセージ・パラメータが、前記メッセージが転送された後に、該メッセージに対して実行される操作を含む、前記（1）記載の方法。

（14）前記2次元配列の前記行見出しとして出現する前記ノード表現が、前記メッセージの転送元の前記ノードを表し、前記2次元配列の前記列見出しとして出現する前記ノード表現が、前記メッセージの転送先の前記ノードを表す、前記（1）記載の方法。

（15）前記2次元配列内の各ノードに対して、前記行見出しとして出現する前記ノード表現の各々と、前記2次元配列内の各ノードに対して、前記列見出しとして出現する前記ノード表現の各々との間の各交点を調査するステップと、前記行見出しと前記列見出しとの前記調査済みの交点の各々が、2つの前記ノード間でメッセージが転送されることを示す指示に関連付けられる前記トークンの1つを含むか否かを判断するステップと、前記調査済みの交点が、2つの前記ノード間でメッセージが転送されることを示す指示に関連付けられる前記トークンの1つを含む場合、前記2次元配列上に前記行見出しとして出現する前記ノード表現により表される前記ノードの各々から、前記2次元配列上に前記列見出しとして出現する前記ノード表現により表される前記ノードの各々

へ、メッセージを転送するステップと、を含む、前記（14）記載の方法。

（16）前記調査済みの交点が、2つの前記ノード間でメッセージが転送されることを示す指示に関連付けられる前記トークンの1つを含む場合に、前記2次元配列上に前記行見出しとして出現する前記ノード表現により表される前記ノードから、前記2次元配列上に前記列見出しとして出現する前記ノード表現により表される前記ノードの各々へ、前記メッセージを転送するステップが、前記トークンに関連付けられる前記メッセージ・パラメータに従い実行される、前記（15）記載の方法。

（17）前記2次元配列の前記列見出しとして出現する前記ノード表現が、前記メッセージの転送元の前記ノードを表し、前記2次元配列の前記行見出しとして出現する前記ノード表現が、前記メッセージの転送先の前記ノードを表す、前記（1）記載の方法。

（18）前記2次元配列内の各ノードに対して、前記列見出しとして出現する前記ノード表現の各々と、前記2次元配列内の各ノードに対して、前記行見出しとして出現する前記ノード表現の各々との間の各交点を調査するステップと、前記行見出しと前記列見出しとの前記調査済みの交点の各々が、2つの前記ノード間でメッセージが転送されることを示す指示に関連付けられる前記トークンの1つを含むか否かを判断するステップと、前記調査済みの交点が2つの前記ノード間でメッセージが転送されることを示す指示に関連付けられる前記トークンの1つを含む場合、前記2次元配列上に前記列見出しとして出現する前記ノード表現により表される前記ノードの各々から、前記2次元配列上に前記行見出しとして出現する前記ノード表現により表される前記ノードの各々へ、メッセージを転送するステップと、を含む、前記（17）記載の方法。

（19）前記調査済みの交点が、2つの前記ノード間でメッセージが転送されることを示す指示に関連付けられる前記トークンの1つを含む場合に、前記2次元配列上に前記列見出しとして出現する前記ノード表現により表される前記ノードから、前記2次元配列上に前記行見出しとして出現する前記ノード表現により表される前記ノードの各々へ、前記メッセージを転送するステップが、前記トークンに関連付けられる前記メッセージ・パラメータに従い実行される、前記（18）記載の方法。

（20）前記2次元配列の連続バージョンが構成され、前記2次元配列の前記連続バージョンの各々が、異なる前記ノード表現の交点における前記ノード間で前記メッセージが転送されることを示す指示に関連付けられるトークンを含むことができ、前記ネットワークを通じる前記ノード間の前記メッセージの転送のために、連続する潜在的に異なる通信パターンを実現するものにおいて、前記2次元配列の前記連続バージョンを構成し、3次元配列を形成するステップであって、前記第3次元配列が前記2次元配列の前記連続バージョンを含む、構成ステップと、トークンを前記2次元配列の前記連続バージョンの各々内の前記ノード表現の交点に配置するステップであって、前記トークンが、前記交差ノード表現により表される前記ノード間でメッセージが転送されるか否かを示す指示に関連付けられ、前記メッセージが転送される場合、前記トークンが、前記ノード間で転送される前記メッセージのメッセージ・パラメータに関連付けられ、こうして生成された前記配列の各々が、前記ネットワークを通じる前記ノード間での前記メッセージの転送のための通信パターンを実現する、トークンの配置ステップと、前記トークンの各々を前記指示及び前記メッセージ・パラメータに関連付けるステップと、前記2次元配列の前記連続バージョンの各々の前記ノード表現の交点における前記トークンが、前記メッセージが転送されることを示す指示に関連付けられる場合、前記トークンを有する前記ノード間で前記メッセージを連続的に転送するステップであって、前記メッセージの転送が、前記トークンに関連付けられる前記メッセージ・パラメータにより定義される、メッセージの転送ステップと、を含む、前記（1）記載の方法。

（21）2つの前記ノード間で前記メッセージが転送されない場合、それらの前記ノード表現の交点の前記トークンが0である、前記（20）記載の方法。

（22）前記2次元配列の前記連続バージョンの各々が、コンピュータ・プログラムにより生成される、前記（20）記載の方法。

（23）前記2次元配列の前記連続バージョンの各々の前記行見出しとして出現する前記ノード表現が、前記メッセージの転送元の前記ノードを表し、前記2次元配列の前記連続バージョンの各々の前記列見出しとして出現する前記ノード表現が、前記メッセージの転送先の前記ノードを表す、前記（20）記載の方法。

（24）前記メッセージを転送するステップが、前記2次元配列の前記連続バージョン上の前記行見出しにより定義される前記ノード表現の各々と、前記2次元配列の前記連続バージョン上の前記列見出しにより定義される前記ノード表現の各々との交点を調査するステップと、前記ノード表現の前記調査済みの交点の各々に対して、前記メッセージが転送されることを示す指示に関連付けられる前記トークンの1つが、前記交点に含まれるか否かを判断するステップと、前記調査済みの交点が、2つの前記ノード間でメッセージが転送されることを示す指示に関連付けられる前記トークンの1つを含む場合、前記2次元配列の前記連続バージョン上に前記行見出しとして出現する前記ノード表現により表される前記ノードの各々から、前記2次元配列の前記連続バージョン上に前記列見出しとして出現する前記ノード表現により表される前記ノードの各々へ、前記メッセージを転送するステップと、を含む、前記（23）記載の方法。

（25）前記調査済みの交点が、2つの前記ノード間でメッセージが転送されることを示す指示に関連付けられる前記トークンの1つを含む場合に、前記2次元配列の前記連続バージョン上に前記行見出しとして出現する前記ノード表現により表される前記ノードから、前記2次元配列の前記連続バージョン上に前記列見出しとして出現する前記ノード表現により表される前記ノードの各々へ、前記メッセージを転送するステップが、前記トークンに関連付けられる前記メッセージ・パラメータに従い実行される、前記（24）記載の方法。

（26）前記2次元配列の前記連続バージョンの各々の前記列見出しとして出現する前記ノード表現が、前記メッセージの転送元の前記ノードを表し、前記2次元配列の前記連続バージョンの前記行見出しとして出現する前記ノード表現が、前記メッセージの転送先の前記ノードを表す、前記（20）記載の方法。

（27）前記メッセージを転送するステップが、前記2次元配列の前記連続バージョン上の前記列見出しにより定義される前記ノード表現の各々と、前記2次元配列の前記連続バージョン上の前記行見出しにより定義される前記ノード表現の各々との交点を調査するステップと、前記ノード表現の前記調査済みの交点の各々に対して、前記メッセージが転送されることを示す指示に関連付けられる前記トークンの1つが、前記交点に含まれるか否かを判断するステップと、前記調査済みの交点が、2つの前記ノード間でメッセージが転送されることを示す指示に関連付けられる前記トークンの1つを含む場合、前記2次元配列の前記連続バージョン上に前記列見出しとして出現する前記ノード表現により表される前記ノードの各々から、前記2次元配列の前記連続バージョン上に前記行見出しとして出現する前記ノード表現により表される前記ノードの各々へ、前記メッセージを転送するステップと、を含む、前記（26）記載の方法。

（28）前記調査済みの交点が、2つの前記ノード間でメッセージが転送されることを示す指示に関連付けられる前記トークンの1つを含む場合に、前記2次元配列の前記連続バージョン上に前記列見出しとして出現する前記ノード表現により表される前記ノードから、前記2次元配列の前記連続バージョン上に前記行見出しとして出現する前記ノード表現により表される前記ノードの各々へ、前記メッセージを転送するステップが、前記トークンに関連付けられる前記メッセージ・パラメータに従い実行される、前記（27）記載の方法。

（29）前記2次元配列の前記連続バージョンに対する前記メッセージ転送のために、前記ノードの各々が同一の命令セットを実行する、前記（20）記載の方法。

【図面の簡単な説明】

【図1】交互のノードも接続される、8つの処理ノードに対する環状通信を示す図である。

【図2】20の処理ノードの場合の2次元トーラス通信トポロジを示す図である。

【図3】5つの処理ノードの場合の全組み合わせ通信トポロジを示す図である。

【図4】7つの処理ノードを含む並列処理システムにおいて、単純な環状通信パターンまたは単純な操作パターンを実現する典型的な生起グラフを示す図である。

【図5】ホスト・ノード・メモリを通じる、トークンとメッセージ転送指示及びメッセージ転送パラメータとの関連付けを示す図である。

【図6】トークンとメッセージ指示及びメッセージ・パラメータとの関連付けを示すフロー図である。

【図7】典型的な生起グラフの交点の調査、及びノード間でのメッセージの転送のフロー図である。

【図8】複数ノード・ネットワークの独自の通信パターンを含む生起グラフの連続バージョンを含む3次元配列を示す図である。

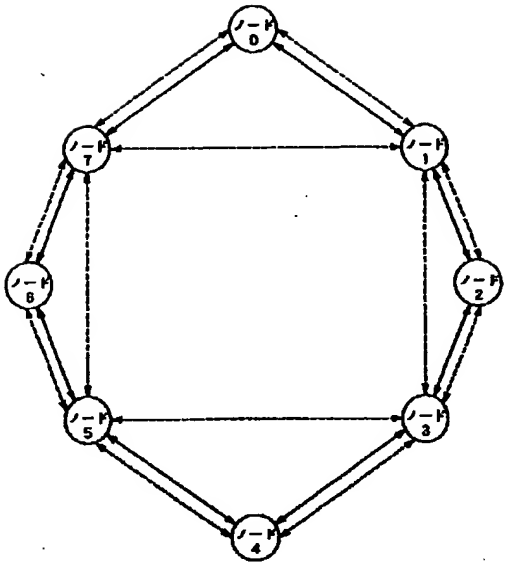
【図9】トークンと非通信操作のためのメッセージ指示及びメッセージ・パラメータとの関連付けを示すフロー図である。

【図10】典型的な生起グラフの交点の調査、及びそのノードにおける非通信操作の実行のフロー図である。

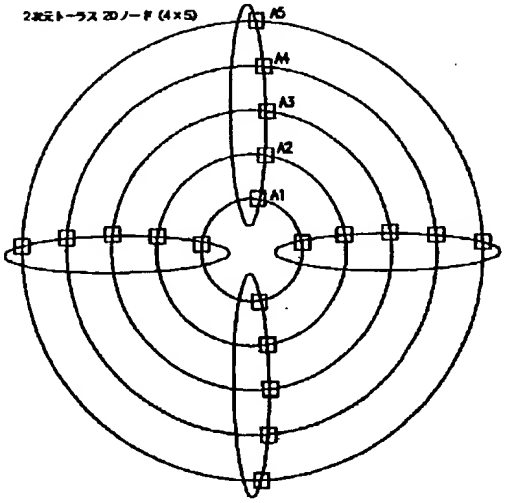
【符号の説明】

400　生起グラフ
401　送信ノード
402　受信ノード
405、405　ベクトル
501　メモリ
502　メモリ位置
505　ホスト・コンピュータ
800　3次元配列
801、802、803、804、805、806、807　2次元生起グラフの連続バージョン

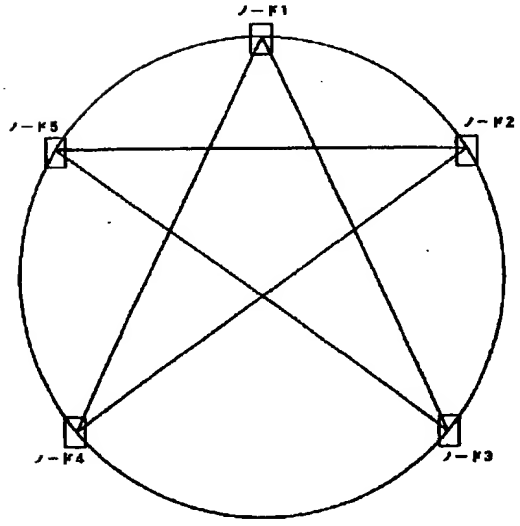
【図1】


ノード0
ノード1
ノード2
ノード3
ノード4
ノード5
ノード6
ノード7

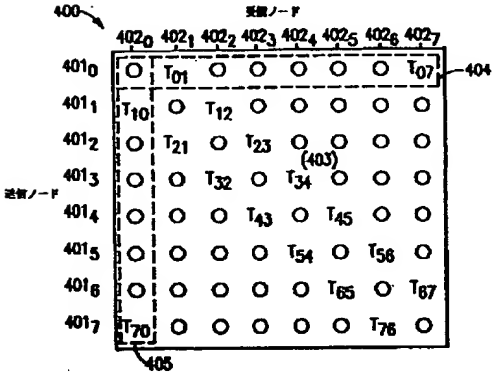
【図2】


2次元トーラス 20ノード (4×5)
A5
A4
A3
A2
A1

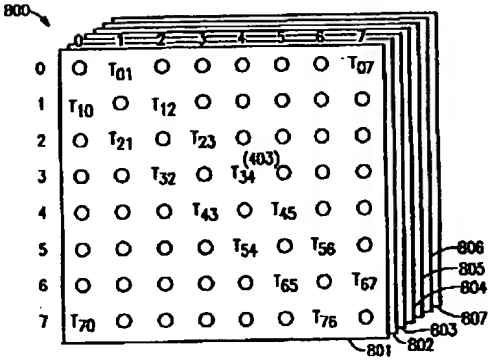
【図3】


ノード1
ノード5
ノード2
ノード4
ノード3

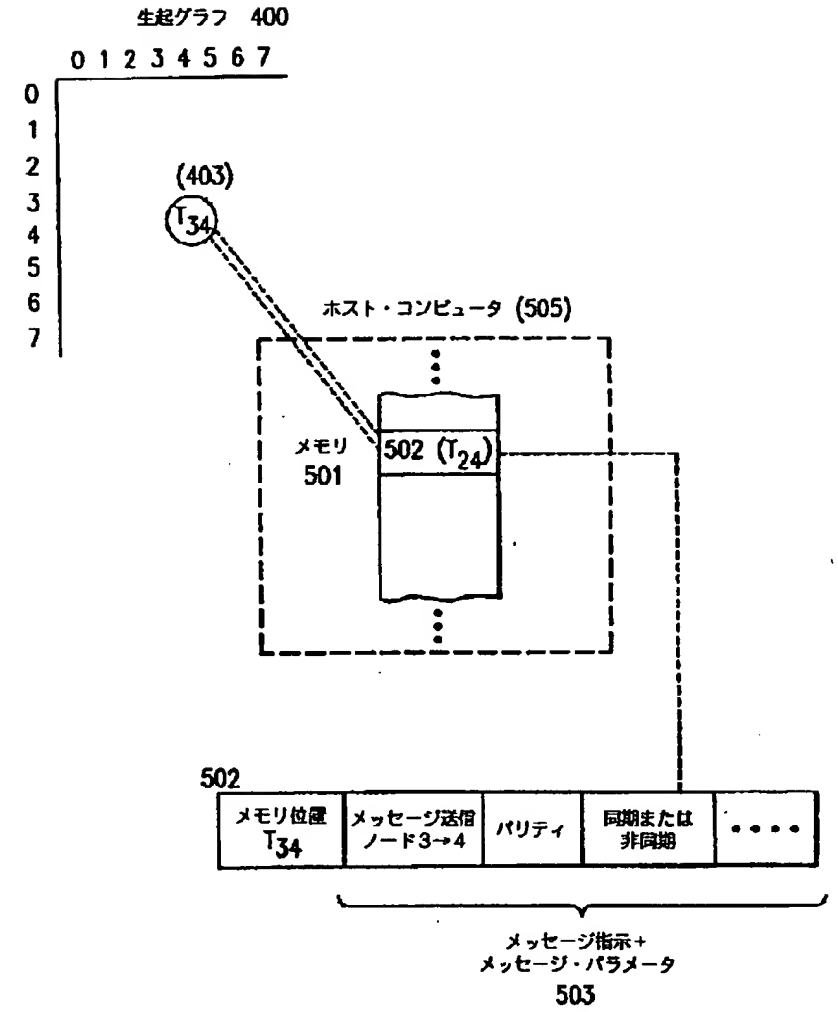
【図4】


400
宛先ノード
送信ノード
404
405
(403)

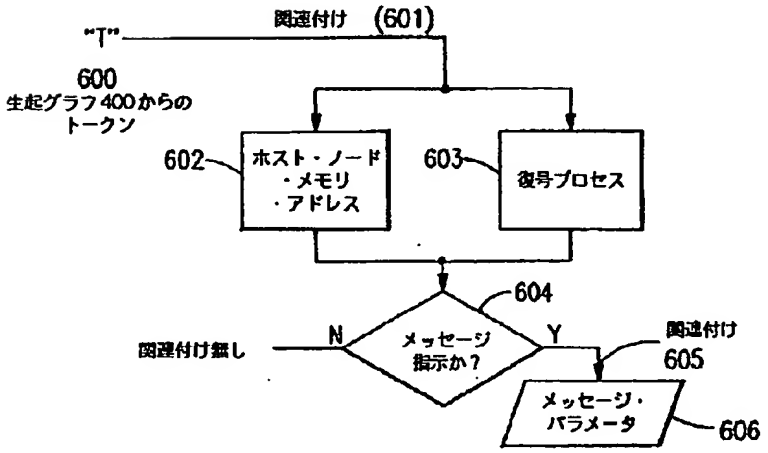
【図8】


800
801
802
803
804
805
806
807
(403)

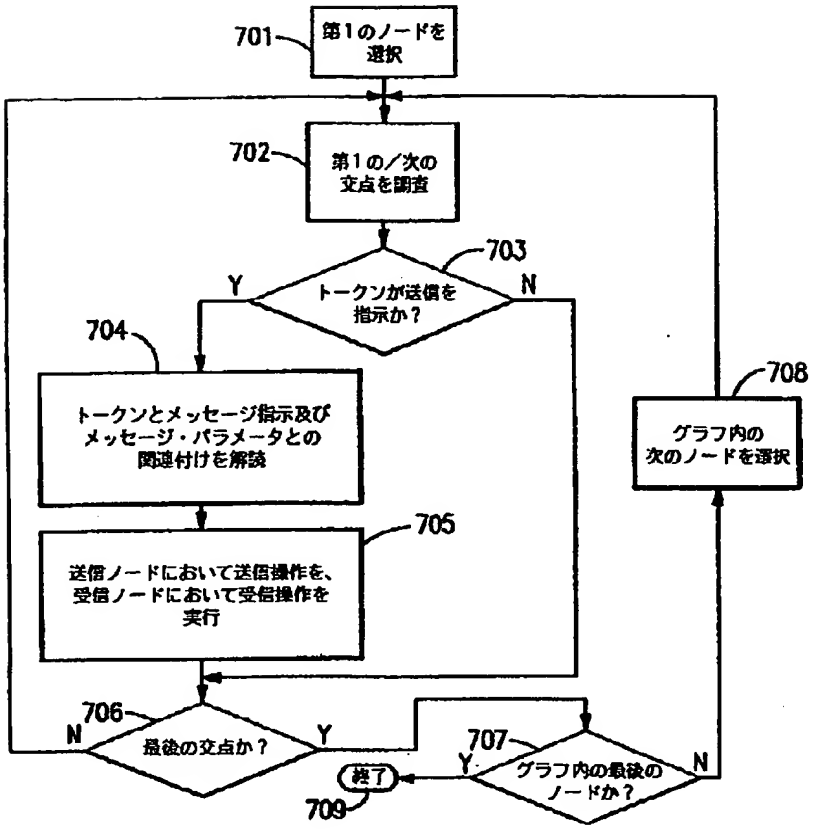
【図5】


生起グラフ 400
0 1 2 3 4 5 6 7
0
1
2
3
4
5
6
7
(403)
T_{34}
ホスト・コンピュータ (505)
メモリ
501
502 (T_{24})
502
メモリ位置
T_{34}
メッセージ送信
ノード3→4
パリティ
同期または
非同期
メッセージ指示+
メッセージ・パラメータ
503

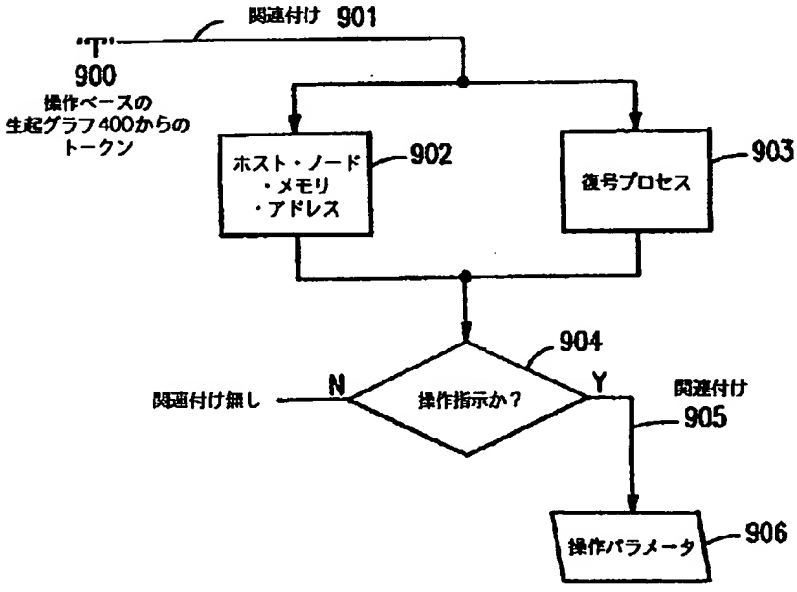
【図6】


"T"
関連付け (601)
600
生起グラフ400からの
トークン
602
ホスト・ノード
・メモリ
・アドレス
603
復号プロセス
604
関連付け無し
N
メッセージ
指示か?
Y
関連付け
605
メッセージ・
パラメータ
606

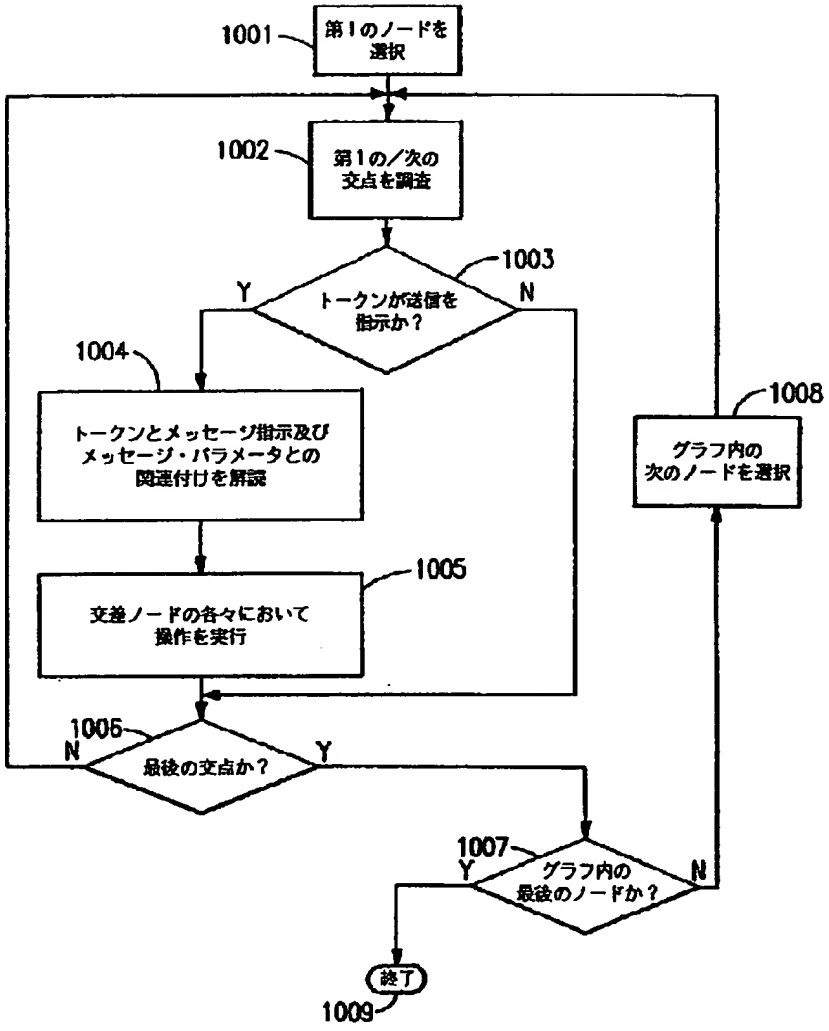
【図7】


701
第1のノードを選択
702
第1の/次の交点を調査
703
トークンが送信を指示か？
Y
N
704
トークンとメッセージ指示及びメッセージ・パラメータとの関連付けを解読
705
送信ノードにおいて送信操作を、受信ノードにおいて受信操作を実行
706
最後の交点か？
N
Y
707
グラフ内の最後のノードか？
Y
N
708
グラフ内の次のノードを選択
終了
709

【図9】


関連付け 901
"T"
900
操作ベースの生起グラフ400からのトークン
ホスト・ノード・メモリ・アドレス
902
復号プロセス
903
904
関連付け無し
N
操作指示か？
Y
関連付け
905
操作パラメータ
906

【図10】



フロントページの続き

(72)発明者　アンドリュー・ワック
アメリカ合衆国12590、ニューヨーク州ワッピンガーズ・フォールズ、タウン・ビュー・ドライブ　53